

---

# A PRACTICAL APPROACH TO USING MULTIPLE-RACE RESPONSE DATA: A BRIDGING METHOD FOR PUBLIC-USE MICRODATA\*

CAROLYN A. LIEBLER AND ANDREW HALPERN-MANNERS

*Revised federal policies require that multiple-race responses be allowed in all federal data collection efforts, but many researchers find the multitude of race categories and variables very difficult to use. Important comparability issues also interfere with using multiple-race data in analyses of multiple data sets and/or several points in time. These difficulties have, in effect, discouraged the use of the new data on race. We present a practical method for incorporating multiple-race respondents into analyses that use public-use microdata. Our method is a modification of the regression method developed by the National Center for Health Statistics (NCHS), which uses multiple-race respondents' specific combination of races, as well as other individual-level and contextual characteristics, to predict the respondents' preferred single race. In this paper we (1) apply the NCHS-generated regression coefficients to public-use microdata with limited geographic information; and (2) provide a downloadable computer program with which researchers can apply this practical and preferable method for including multiple-race respondents in a wide variety of analyses.*

**R**ace is a contextual, contingent, complicated, and life-directing social construct (e.g., Cornell and Hartmann 2007; Harris and Sim 2002; Root 1996). In the late 1990s, American federal policies for collecting data on race changed to better reflect this relatively recent understanding. The revised policies require that multiple-race responses be allowed in all federal data collection efforts and encourage data creators to provide as much detail as possible about their respondents' race reports (Office of Management and Budget [OMB] 1997, 2000). The change in requirements affected Census 2000, and many other data collection efforts have followed suit. As witnesses to this great change in the way race is recorded, contemporary researchers have the opportunity to describe our complex social world more accurately and with more nuance.

In reality, however, many researchers find the multitude of race categories and variables very difficult to use, discuss, and interpret (Harrison 2002; Snipp 2003). Definitional changes and inconsistent classification schemes interfere with the calculation of statistics that rely on multiple data sets that measure and record race differently. In contrast to the new federal guidelines, for example, most state governments allow only one race response when collecting vital statistics on births, deaths, and marriages. Researchers seeking to assess change over time encounter similar difficulties when working with historically incomparable measures.

In this paper, we offer a practical tool for dealing with this problem: a “bridge” between the prior single-response race question format and the new “mark all that apply” system. The bridging tool we present allows a researcher to recode multiple-race responses

---

\*Carolyn A. Liebler and Andrew Halpern-Manners, Department of Sociology and Minnesota Population Center, University of Minnesota. Address comments to Carolyn Liebler, Department of Sociology, 267 19th Ave. S., Minneapolis, MN 55455; E-mail: [lieblier@umn.edu](mailto:lieblier@umn.edu). The first author began this project while she was funded by “IPUMS-Redesign” (NIH GRANT R01-HD043392), Steven Ruggles, Principle Investigator. We presented a draft of this work at the 2007 annual meetings of the Population Association of America, New York. We thank J. Trent Alexander, Douglas Hartmann, Elaine M. Hernandez, Deborah D. Ingram, C. Matthew Snipp, and John Robert Warren for their helpful feedback, and the Minnesota Population Center for its invaluable research support.

into single-race categories by calculating the predicted probability that the respondent would have chosen a particular race response if asked to choose only one. Unlike existing methods, our “modified regression method” can be used in conjunction with most public-use data sets to calculate changes across time in the same data series, to compare data sets with different race question formats, or simply to code “race” in a data set with a very high number of race categories.

## BACKGROUND

### Three Problematic Approaches

To avoid the complexity inherent in data with multiple-race responses, researchers typically use one of three approaches. Many scholars exclude multiracial people from their study in order to simplify their analyses. When doing so, researchers assume that multiracial people are, on average, the same as single-race people, such that excluding them does not bias the sample or any substantive conclusions. In most cases, this is not true. Multiracial respondents are distinctive both qualitatively (e.g., Liebler 2001; Rockquemore and Brunsma 2002; Root 1996) and quantitatively. As shown in Table 1, multiracial American Indians are more likely to live in urban areas and have higher incomes than other American Indians. Analogous significant differences exist among Pacific Islanders, Asians, and blacks.

Other researchers group all multiracial persons into a single residual category. This strategy creates its own problems. When diverse populations are collapsed into a single group, the results for this residual group cannot be interpreted. All cultural relevance is lost, as is potential understanding of the component groups’ experiences (Burhansstipanov and Satter 2000; National Committee on Vital and Health Statistics 2005). This strategy is equivalent to excluding minority groups from analyses for the convenience of the researcher, which is not consistent with current ethical standards (e.g., LaVeist 1994; National Institutes of Health 2001).

A third and less frequently used technique involves grouping multiracials with each of their related monoracial groups (hereafter referred to as the *all-inclusive method*). For example, this approach would include information about an American Indian–black person in the “black alone or in combination” category *and* in the “American Indian alone or in combination” category. In addition to complicating cross-time comparisons, this method is not entirely intuitive, primarily because it yields subtotals that do not sum to the total number of cases.

### Two Types of Bridging Methods

As alternatives to these problematic approaches, federal agencies and academics have developed ways to “bridge” multiple-race responses into single-race categories (see Tucker, Miller, and Parker 2002). Many of these strategies recode each multiracial response into a single-race category based on a predetermined assignment rule using *whole assignment*. If the assignment rule prioritizes the racial group with the largest population, for example, the researcher would allocate all of the “black and Japanese” responses to the “black” category. Deterministic whole-assignment methods are straightforward to apply and explain but are limited in two respects: (1) the choice of assignment rule can have a powerful effect on the results (e.g., Parker and Makuc 2002); and (2) they introduce measurement error by retaining incomplete information about respondents’ reported races.

More refined bridging techniques use *fractional assignment*. Most fractional assignment methods apply a predetermined fractional weight to each multiple-race response. The “equal fractional assignment” method, for instance, requires that each “Chinese and white” response be recoded into two categories, “Chinese” and “white,” each with a weight of 0.5. Virtually all fractional assignment methods provide improved approximation to past racial distributions (Allen and Turner 2001; Grieco 2002; Heck et al. 2003; Lee 2001; Parker and

Table 1. Racial Group Differences on Selected Economic, Demographic, and Geographic Characteristics in 2000

	American Indian and/or Alaska Native		Pacific Islander		Asian		Black or African American	
	Multiracial <sup>a</sup>	Monoracial <sup>b</sup>	Multiracial <sup>a</sup>	Monoracial <sup>b</sup>	Multiracial <sup>a</sup>	Monoracial <sup>b</sup>	Multiracial <sup>a</sup>	Monoracial <sup>b</sup>
Mean Personal Income (ages 15+)	21,522***	17,879	21,121***	20,338	22,882***	27,449	20,851***	28,722
Percentage Living in a Rural Area	18.0***	33.7						
Percentage Living in Hawaii			34.1***	29.7				
Percentage Under Age 18	31.6	33.1	39.0***	31.3	44.7***	23.9	36.3***	24.4

Source: Integrated Public Use Microdata Series (IPUMS), Census 2000, 5% sample.

<sup>a</sup>Includes persons who reported, for example, American Indian/Alaska Native as well as other racial group(s).

<sup>b</sup>Includes persons who only reported one race.

\*\*\*Difference between multiracial and monoracial subgroups is statistically significant at  $p < .001$ .

Makuc 2002). However, using predetermined fractions remains problematic. Deterministic assignment rules, whether fractional or whole, ignore the substantial variation in single-race response patterns found within each multiracial population (Parker et al. 2004)—a limitation that biases the bridged results, especially with respect to smaller groups (Grieco 2002; Heck et al. 2003; Lee 2001; Mays et al. 2003; Parker and Makuc 2002; Schenker and Parker 2003).

### **The NCHS Regression Method**

In hopes of developing a better bridging method, many federal researchers and policy makers have focused on data collected in the National Health Interview Survey (NHIS) (Ingram et al. 2003; OMB 2000; Parker et al. 2004; Schenker and Parker 2003). The NHIS has allowed multiple-race responses for decades, asking each multiple-race respondent a follow-up question to identify the single race that best describes him- or herself. With this information, researchers at the National Center for Health Statistics (NCHS) used multivariate methods to predict the single-race response preferred by each multiracial NHIS respondent in 1997–2002 (Ingram et al. 2003). Sensitivity analyses indicate that this allocation approach creates significantly less bias and has greater predictive power than other bridging methods (Schenker and Parker 2003).

The NCHS researchers' multivariate models included age, sex, and Hispanic origin as individual-level covariates. These variables are known to influence individuals' responses when asked about their race (Liebler 2004; Rockquemore 2002; Rodríguez 2000; Waters 1999) in ways that vary among multiracial groups (Kana'iaupuni and Liebler 2005). Appropriately, the NCHS estimated separate models for each multiracial group. They also used covariates to control for the contextual factors of region, urbanicity, and the racial composition of the area, all of which are tied to race responses (Kana'iaupuni and Liebler 2005; Liebler 2004; Xie and Goyette 1998). In collaboration with the Census Bureau, the NCHS researchers calculated the latter two variables using county-level geographic information from private census data files.

Despite the high quality of the NCHS regression method, researchers attempting to apply the resulting regression coefficients to other data sets face substantial challenges; the NCHS method is impossible to fully implement without detailed information about individuals' geographic location. In this paper, we present a modified regression method for bridging and include computer code that researchers can apply to public-use microdata with state-level geographic identifiers. In the sections that follow, we describe the specific compromises necessary for applying our bridging method, introduce the usage of bridged race data, and document the small differences in estimates that result from using state-level measures of racial context.

### **METHODS**

We have made modifications to the NCHS regression method in order to apply it to public-use microdata with limited geographic information. We label our method the "modified regression method" and provide it in a downloadable STATA program at the following Web address: [http://usa.ipums.org/usa/volii/race\\_bridge\\_stata\\_program.txt](http://usa.ipums.org/usa/volii/race_bridge_stata_program.txt).

When bridging complex race data using the modified regression method, the researcher provides the following individual-level information: specific race responses, age, sex, Hispanic origin, and state. The bridging program identifies the individual's region and provides imputed urbanization level and racial diversity information based on the state of residence. The program then applies the NCHS-provided regression coefficients for these variables to calculate the probability of each single-race response for each multiple-race respondent in the data set. These probabilities are then converted into a set of bridged race variables that analysts can use for either fractional assignment or whole assignment of multiracial respondents to single-race categories. For some exceptionally well-used data—specifically,

Census 2000 and the American Community Survey (ACS)—the bridged race variables described in this article can be downloaded directly rather than calculated.<sup>1</sup>

### Compressing Multiple-Race Responses

In order to work with a full set of possible multiple-race responses in the context of potentially hundreds of ways to mark multiple races, the NCHS team compressed the multiracial groups into 11 multiple-race categories (hereafter referred to as the *modified race data format*), representing all of the possible combinations of (1) American Indian and/or Alaska Native (AIAN or American Indian); (2) Asian and/or Pacific Islander (API or Asian/Pacific Islander); (3) black or African American (B or black); and (4) white (W). The NCHS calculated regression coefficients separately for most of these 11 multiple-race groups, thus allowing for differences between the major group combinations in the strength and direction of predictors.

Two key comparability issues arise from this coding scheme. First, the modified race data format combines Asians and Pacific Islanders. This is done because Asians and Pacific Islanders were usually tallied together in the past, and bridging is intended to mimic the past. Thus, individuals who mark several Asian groups, or who mark an Asian and a Pacific Islander group, are not considered multiracial under this categorization scheme. Second, some data sets include respondents who mark “some other race” (SOR) and no additional race(s). Most of these individuals are Hispanic/Latino. Because federal guidelines do not recognize Hispanic/Latino as a race (OMB 1997), the modified race data format recodes respondents who marked SOR as well as another race (or races) to their non-SOR response(s). We elected not to reassign the races of SOR respondents, primarily because Hispanics of “some other race” often prefer not to identify with any of the available race categories (e.g., Rodríguez 1992, 2000).<sup>2</sup>

### Limited Geographic Information

Most public-use microdata contain detailed information that must be kept confidential. A common way of reducing the risk of a breach in confidentiality is to restrict the amount of geographic information available. In the public version of the Census 2000 data (5% sample), for example, the lowest level of geographic information available is the person’s Public Use Microdata Area—a census-defined area with a minimum of 100,000 people. In other data (e.g., the Current Population Survey [CPS] and the 2000–2004 ACS), the lowest level of available geography is the state. In calculating the original regression coefficients, the NCHS team had access to geographic detail down to the county level and incorporated this into their analysis. They used this information in two ways for their regression predictions: (1) they calculated the racial composition of the respondent’s county using the internal Census Bureau files; and (2) they coded the urbanization level of each respondent’s local area. Because they used restricted-use county-level data, the geographic aspects of the NCHS regression method cannot be replicated using most publicly available data—compromises must be made.

**Recalculating racial composition of the area.** Using the modified race data format we described, the NCHS regression method measures the racial composition of local areas using four variables: percent American Indian in the county, percent Asian/Pacific Islander in the county, percent black in the county, and percent multiple-race in the county.

---

1. See the variable RACESING at <http://usa.ipums.org/usa>. Note that Census 2000 1% and 5% Microdata provide more detailed geographic information. We used the greatest available level of geographic detail for the calculations disseminated via the Integrated Public Use Microdata Series (IPUMS) Web site.

2. This is a departure from the NCHS regression method. The NCHS assigned single-race SOR respondents to a single- or multiple-race response using hot-deck imputation. Respondents who were allocated to the latter group were subsequently bridged back to their most likely single-race group. Thus the allocation of SOR responses enters into the NCHS’ original regression calculations but not the modified regression results presented here.



When creating parallel variables using state-level information, it is important to code race responses in the modified race data format before calculating the percent of each racial group in the area. Complete replication of the modified race data format is not always possible, however, because this format requires that “some other race” responses be allocated as we discussed earlier. To maximize comparability with the NCHS method, we used the state-level Census 2000 modified race data (provided in Ingram et al. 2003) to calculate the state-level racial composition that is used in our bridging program.

**Recalculating urbanization level.** More consequential compromises must be made when working to characterize the urbanization level of each respondent’s local area, given limited geographic information. For their regression, the NCHS researchers measured urbanization level using four categories: large urban, large suburban, medium/small metropolitan, and nonmetropolitan.<sup>3</sup> To create these categories for respondents in publicly available microdata, the researcher needs the following information: (1) whether the person lives in a “large” city; and (2) if so, whether he/she lives in the urban part or the suburban part of the city, or (3) if not, whether he/she lives in a smaller city as opposed to an area not defined as a city. This information is rarely available in public-use data sets.

For the modified regression method, we calculated the proportion of individuals in the state living in each of the four types of places using full-count data from Census 2000 (SF1, Table GCT-PH1).<sup>4</sup> Then we assigned each person in the state the same value for each of the four urbanization level indicators. For example, 29.59% of people in Minnesota in 2000 (1,456,119 of 4,919,479) lived in nonmetropolitan areas. Thus every resident of Minnesota is assigned a value of .2959 for their “nonmetropolitan” variable. This geographic restriction forces the questionable—but necessary—assumption that all residents are equally likely to live in a nonmetropolitan area; in truth, multiracial individuals are geographically concentrated in complex ways (Farley 2002; Jones and Symens Smith 2001).<sup>5</sup>

### Selecting a Data Set

We used data from Census 2000 in the imputations and calculations above, as well as in the bridging program provided electronically. We utilized this data source, rather than a more recent one, for three reasons. First, Census 2000 is widely analyzed, and it is appropriate to build race bridges using a contemporaneous data set. Second, the Census 2000 full-count population data (SF1) provide detailed information that is not available elsewhere. Substituting more modern data for some of the imputations and calculations but not others would muddle the situation further. Third, most noncensus data sets to which one might apply the modified regression method are at most only a few years younger than Census 2000. Bridging proportions such as those calculated by the NCHS change over time (Schenker and Parker 2003) and should be recalculated as often as data availability permits. Until the results of the 2010 census are released, however, researchers applying our modified regression method to publicly available microdata are working with the best available resource for coding complex race responses into usable and meaningful categories.

### APPLICATION OF THE METHOD

The bridging equations in the modified regression method use individual-level and contextual information about multiple-race respondents to assign each multiracial person four weights. Each weight represents the predicted probability that the person would have

3. In this context, “large” is defined as a city of 1 million or more population (see Eberhardt et al. 2001: 78–80).

4. We coded residence in a “large urban area” as being equivalent to living inside of the area’s “central city.”

5. To improve upon this assumption, a researcher would use information about which types of multiracial individuals are likely to live in which cities, including whether they are likely to live in the central city areas and how much their distribution differs from that of the single-race population. This refinement is beyond the scope of this article.

reported that particular single race (AIAN, API, B, or W) if asked to choose only one. These weights can be used for fractional assignment; alternatively, the individuals can be wholly assigned to their most heavily weighted single race. In this section, we provide a brief overview of the usage of our bridging program and provide examples of how to use the variables it creates.

As discussed at the beginning of the electronically available bridging program, the data in need of bridging must be individual-level data. Each multiple-race individual's race, age, sex, Hispanic origin (yes/no), and state of residence must be included. The bridging program creates five variables. The first four variables—AIPROB, APIPROB, BPROB, and WPROB—represent the probability that the individual would have reported each of the four single races. Again, these can be interpreted as weights and can be used for fractional assignment. The fifth variable, ONERACE, provides the single race that the person is most likely to have reported; that is, it indicates which of the first four variables has the highest value. ONERACE is the variable to use if the researcher prefers whole assignment.

### Practical Application of Fractional Assignment

To incorporate single-race respondents into this race coding scheme, researchers using the four variables AIPROB, APIPROB, BPROB, and WPROB will need to assign single-race individuals a value of 1 on the relevant variable. Researchers may also need to create a variable indicating a single-race “some other race” response (included here as SORPROB) if the source data include this category. Using these variables in multivariate analyses simply requires that the researcher include this set of continuous variables in the model instead of the more familiar strategy of measuring race through dummy variables.

The advantage of using fractional assignment over whole assignment is best shown through example. The data given in Table 2 represent three individuals. Person 1 is a single-race white person. Person 2 is an American Indian and white person whose personal characteristics and locational context imply that he is likely to have reported white single race if asked to choose; however, he also has a nonzero probability of choosing American Indian/Alaska Native. Person 3 is a different American Indian and white person whose characteristics and context make it more probable that she would have reported American Indian if asked to choose (AIPROB = .546). The final column (ONERACE) provides the single race that is assigned to each person using whole assignment. Note that unlike whole assignment, fractional assignment retains information about each of the respondents' self-reported races—an identification that includes two or more groups on purpose. As we show below, this enhanced sensitivity allows the bridge to provide a better approximation of the previous race question format.

### Practical Application of Whole Assignment

Despite having desirable properties, the fractional assignment of variables will be seen by some researchers as too cumbersome to use effectively, especially if some race groups are to be excluded from the study sample. For example, if a study using fractional bridged race were to focus only on the experiences of blacks and whites, a biracial black and American Indian person would be only partially included in the study. For these situations, whole assignment bridging is preferable. The ONERACE variable generated by the bridging program assigns the whole multiracial person's case to a single race category, based on

**Table 2. Fractional Assignment and Whole Assignment Variables for Three Example Cases**

	AIPROB	APIPROB	BPROB	SORPROB	WPROB	ONERACE
Person 1	0	0	0	0	1	white
Person 2	.101	0	0	0	.899	white
Person 3	.546	0	0	0	.454	AIAN

the multiracial respondent's personal characteristics and context. If ONERACE is to be used in a multivariate analysis, we recommend also including an indicator of whether the respondent reported multiple races.

Whole assignment approximates fractional assignment in cases, such as Person 2 in Table 2, in which the highest response probability is close to 1. In cases such as Person 3, however, more error is introduced by the use of the whole assignment method. Groups whose probability of assignment to each single-race group (i.e., their weight) is typically near 0.5 are especially affected by the decision to use whole assignment.

Because of increased precision and decreased bias, we favor the fractional allocation method represented in the probability variables AIPROB, APIPROB, BPROB, and WPROB. However, the whole assignment method provided here is preferable to all of the other whole assignment methods discussed previously because it incorporates relevant information about the individual into the prediction of his or her most likely single race. Both recoding methods given here—fractional assignment and whole assignment—allow for meaningful variation in assigned race codes within multiracial populations while providing a practical number of race categories.

### A Cautionary Note

At the individual level, bridged race should be treated with caution. A person's bridged race is a point estimate with a high standard error because the independent variables in the bridging regression explain only a small part of the variance captured in the complex race question (Parker et al. 2004; Schenker and Parker 2003). Bridged estimates were developed with the intention of generating aggregate-level statistics, so that errors at the individual level would average out. One consequence of high error at the individual level is that bridged race is not appropriate for use as a dependent variable. This is especially true if predictors include age, sex, Hispanic origin, and/or racial context measures. Bridged race variables may, however, be used as independent variables—even in combination with the demographic and context variables used to create bridged race. Single-race respondents whose race responses have not been bridged abate any collinearity that might otherwise exist.

## RESULTS

To evaluate the modified regression method we make two comparisons. First, we compare our fractional assignment weights, calculated using state-level geographic information, with those that the NCHS calculated using county-level diversity data and nonimputed urbanization data. These comparisons are intended to reveal the location and extent of biases introduced when applying state-level measures of context instead of more-detailed indicators. Second, we illustrate the improved cross-time comparability in estimates by comparing our whole assignment and fractional assignment results with results calculated using the all-inclusive method described previously.

### Comparisons With the NCHS Regression Method

The results of the modified regression method—when applied to Census 2000 5% PUMS data using state-level information—compare favorably with the results of the NCHS bridging method, as applied to the NHIS data (Ingram et al. 2003). In Table 3 we compare the mean values for each of the fractional assignment variables (AIPROB, APIPROB, BPROB, and WPROB) within each of the 11 multiple-race categories. For instance, on average, the probability of assignment to American Indian among those who reported American Indian/Alaska Native and black was .186 using restricted-use measures of geographic information and context—an estimate that is only 2.3 percentage points from the estimate produced using the modified regression method (.163).

Applying the NCHS regression coefficients to data with only state-level racial context measures has the most substantial biasing effect on the fractional weights generated for



Table 3. Mean Value of Fractional Assignment Weights, by Multiracial Group and Data Source

Multiple-Race Response	Mean Single-Race Weight							
	American Indian and/or Alaska Native (AIAN)		Asian and/or Pacific Islander (API)		Black or African American (Black)		White	
	Private Data <sup>a</sup>	Public Data <sup>b</sup>	Private Data <sup>a</sup>	Public Data <sup>b</sup>	Private Data <sup>a</sup>	Public Data <sup>b</sup>	Private Data <sup>a</sup>	Public Data <sup>b</sup>
AIAN and API	.404	.363	.596	.637				
AIAN and Black	.186	.163			.814	.838		
AIAN and White	.205	.221					.795	.779
API and Black			.370	.350	.630	.650		
API and White			.327	.401			.673	.599
Black and White					.621	.607	.379	.394
AIAN, API, and Black	.286	.327	.253	.255	.461	.418		
AIAN, API, and White	.024	.023	.043	.084			.933	.893
AIAN, Black, and White	.195	.192			.572	.626	.233	.182
API, Black, and White			.104	.103	.113	.098	.782	.800
AIAN, API, Black, and White	.010	.010	.009	.009	.020	.013	.960	.967

<sup>a</sup>"Private Data" columns represent NCHS regression estimates using restricted-use county-level data from four years of the National Health Interview Survey (Ingram et al. 2003: table 9).

<sup>b</sup>"Public Data" columns represent the modified regression method applied to the Census 2000 5% microdata (weighted) using state-level geography only. These columns represent the fractional assignment weights (AIPROB, APIPROB, BPROB, and WPROB) generated by the program provided online at [http://usa.ipums.org/usa/voliii/race/bridge\\_stata\\_program.txt](http://usa.ipums.org/usa/voliii/race/bridge_stata_program.txt).

Asian/Pacific Islanders and American Indians. For example, when county racial composition and observed urbanization information is used in the calculation (as by Ingram et al. 2003), multiracial AIAN/API respondents have an average predicted probability of .404 of reporting American Indian/Alaska Native instead of Asian/Pacific Islander, if asked to choose just one. Using publicly available state-level detail about racial diversity and imputing urbanization information decreases the mean of this weight to .363. This artificial difference in means can be fully attributed to the compromises we described earlier. The difference between the NCHS regression method and our modified regression method is relatively large for the AIAN/API group and for the API/W group—a difference in means of .041 and .074, respectively—and is slight for the other nine multiracial groups. Unlike other racial groups, Asian/Pacific Islanders (including those who are multiracial) are especially likely to live in large urbanized areas, while American Indians (including those who are multiracial) are especially likely to live in rural areas. State-level data hide these variations in context.

Note that the differences in the mean values of fractional allocation weights are consequential in subtly different ways, depending on whether the researcher uses whole assignment or fractional assignment. Using fractional assignment and the modified regression method, the researcher will calculate a slightly higher value for AIPROB among API/W and AIAN/API respondents than she would with full geographic information. Using whole assignment and the modified regression method, the researcher would instead assign slightly more multiracial individuals in these groups to the Asian/Pacific Islander category of ONERACE. Overall, differences are minimal and are highlighted here only for completeness.

### Comparisons With Unbridged Estimates

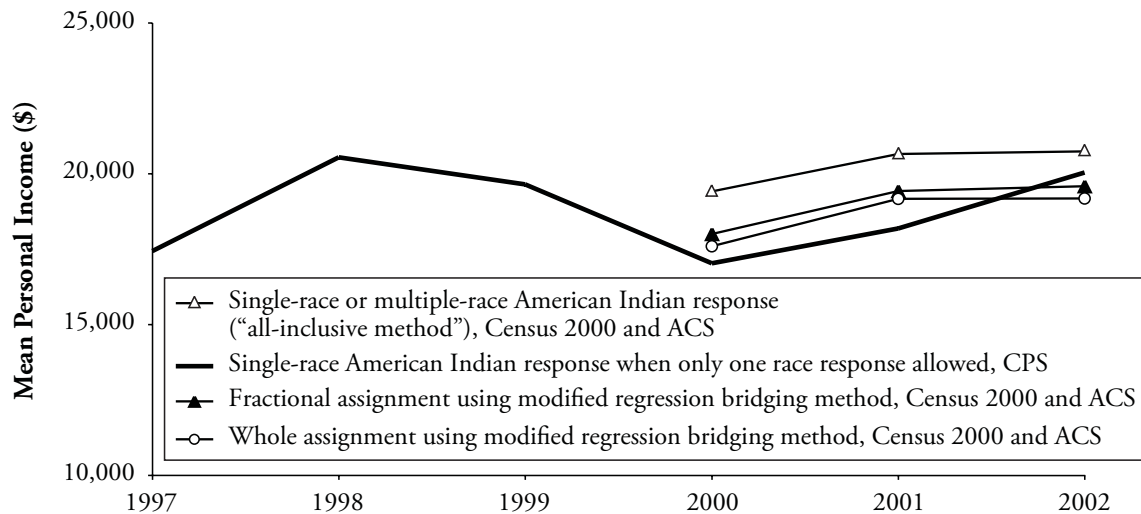
In terms of historical consistency, the modified regression method provides more accurate results than unbridged estimates. The four lines in Figure 1 represent four cross-time estimates of the average personal income of American Indian/Alaska Natives ages 15 and older. Because it used the forced-choice single-race question wording until 2002, changes in the CPS estimates of American Indian income can be seen as real changes. In comparison, Census 2000 and the ACS used the “mark all that apply” multiple-race response system beginning in 2000. In this respect, the differences between the CPS estimates for 2000–2002 and the census/ACS estimates illustrate the distortion that can be introduced when working with race data that span classification regimes.

The results presented in Figure 1 provide a demonstration of the face validity of the modified regression method. We used state-level geographic information in Census 2000 and ACS 2001–2002 public-use microdata to calculate both AIPROB and ONERACE (shown in separate lines). Whereas the income estimates generated using the unbridged all-inclusive method vary by as much as 14% from the CPS benchmark, the modified regression method provides a much closer approximation of the CPS results for each of the years in question. Both the fractional assignment (AIPROB) and the whole assignment (ONERACE) versions of the modified regression method provide high-quality estimates of real temporal changes in a population’s characteristics and minimize the disruption due to changes to the race question.

### DISCUSSION

Researchers in sociology (Snipp 2003), public health (Mays et al. 2003; National Committee on Vital and Health Statistics 2005), and education policy (Renn and Lunceford 2004) have highlighted the need for practical methods for incorporating newly complex race data into analyses that require consistent measures of race. In this paper, we provide and document just such a method. Unlike other race bridging methods, researchers can apply the modified regression method to a wide variety of commonly used and publicly available microdata sets, thereby avoiding the pitfalls of folding all multiracial persons into a single residual

**Figure 1. Mean Personal Income Among American Indians/Alaska Natives Ages 15+, 1997–2002: A Comparison of Two Unbridged Estimates and Two Bridged Estimates**



*Notes:* All estimates are presented in constant 2000 dollars. ACS = American Community Survey (2001, 2002). CPS = Current Population Survey (1997–2002).

category or dropping such cases altogether. While still reflecting respondents' race responses, the modified regression method allows researchers to make relatively accurate cross-time comparisons by retaining historically consistent and substantively meaningful groupings of people. In other words, this method for working with complex multiple-race data is both preferable and practical.

We have presented two ways of applying the modified regression method: fractional assignment and whole assignment. Both approaches use multiracial respondents' key characteristics and contextual information in order to predict each individual's most likely single-race answer to a forced-choice race question. The whole assignment method presented here provides the single race most likely to be reported by each multiracial respondent. The fractional assignment method is slightly more cumbersome to use but represents a more nuanced approach to a complex situation by providing nonzero predicted probabilities (or weights) of each potential single-race response for each multiracial respondent.

The research community stands to benefit in at least three ways from using the modified regression method. First, the method provides improved measurement. Analysts who use this methodologically sound and substantively meaningful approach to generate simplified race variables will avoid problems of bias and/or incomparability inherent in other methods. Second, the method is timely. Although the number of multiracial individuals may seem small now, this number is likely to grow (Goldstein and Morning 2000; Lee 2001; Waters 2000); it is important to implement a good bridging method as early as possible during the transition to new race data so that research done now is not undermined in the future by questions about how multiracial responses were used (or ignored). And finally, the method is practical. The downloadable computer program is straightforward to implement and encourages comparability and consistency between research projects from a variety of disciplines.

Because of the fluidity and context-dependence of race, measuring it at all in a survey remains inherently challenging. No survey question can fully measure the multifaceted and

ever-changing social construct of “race” (Liebler 2001; Nobles 2000; Renn 2004; Rockquemore and Brunnsma 2002; Wallace 2001). In the end, we must caution that, like all bridging methods, the modified regression method represents no more than an educated guess about an unobserved situation. Nevertheless, the multiple-race population exists and is increasing. Suppressing multiracial responses through aggregation or exclusion introduces bias and misrepresents populations; this practice should be avoided when at all possible. By disseminating a sophisticated, practical, and well-documented approach to using complex multiple-race data, we allow analysts to retain much of the meaningful information that can be gathered through a survey question about race—and thus remain sensitive to the complexities of race while fulfilling the need for historical or cross-survey compatibility.

## REFERENCES

- Allen, J.P. and E. Turner. 2001. “Bridging 1990 and 2000 Census Race Data: Fractional Assignment of Multiracial Populations.” *Population Research and Policy Review* 20:513–33.
- Burhansstipanov, L. and D.E. Satter. 2000. “Office of Management and Budget Racial Categories and Implications for American Indians and Alaska Natives.” *American Journal of Public Health* 90(11):1720–23.
- Cornell, S. and D. Hartmann. 2007. *Ethnicity and Race: Making Identities in a Changing World*. 2nd ed. Thousand Oaks, CA: Pine Forge Press.
- Eberhardt, M.S., D.D. Ingram, D.M. Makuc et al. 2001. *Health United States, 2001 With Urban and Rural Health Chartbook*. Hyattsville, MD: National Center for Health Statistics.
- Farley, R. 2002. “Racial Identities in 2000: The Response to the Multiple-Race Response Option.” Pp. 33–61 in *The New Race Question: How the Census Counts Multiracial Individuals*, edited by J. Perlmann and M.C. Waters. New York: Russell Sage Foundation.
- Goldstein, J.R. and A.J. Morning. 2000. “The Multiple-Race Population of the United States: Issues and Estimates.” *Proceedings of the National Academy of Sciences* 97(11):6230–35.
- Grieco, E.M. 2002. “An Evaluation of Bridging Methods Using Race Data From Census 2000.” *Population Research and Policy Review* 21:91–107.
- Harris, D.R. and J.J. Sim. 2002. “Who is Multiracial? Assessing the Complexity of Lived Race.” *American Sociological Review* 67:614–27.
- Harrison, R.J. 2002. “Inadequacies of Multiple-Response Race Data in the Federal Statistical System.” Pp. 137–60 in *The New Race Question: How the Census Counts Multiracial Individuals*, edited by J. Perlmann and M.C. Waters. New York: Russell Sage Foundation.
- Heck, K.E., J.D. Parker, C.J. McKendry, and G.F. Chávez. 2003. “Mind the Gap: Bridge Methods to Allocate Multiple-Race Mothers in Trend Analyses of Birth Certificate Data.” *Maternal and Child Health Journal* 7(1):65–70.
- Ingram, D.D., J.D. Parker, N. Schenker, J.A. Weed, B. Hamilton, E. Arias, and J.H. Madans. 2003. “United States Census 2000 Population With Bridged Race Categories.” *Vital Health Statistics* 2(135). Hyattsville, MD: National Center for Health Statistics.
- Jones, N.A. and A. Symens Smith. 2001. “The Two or More Races. Population: 2000.” Census 2000 Brief C2KBR/01-6. U.S. Census Bureau, Washington, DC.
- Kana’iaupuni, S.M. and C.A. Liebler. 2005. “Pondering Poi Dog: Place and Racial Identification of Multiracial Native Hawaiians.” *Ethnic and Racial Studies* 28:687–721.
- LaVeist, T.A. 1994. “Beyond Dummy Variables and Sample Selection: What Health Services Researchers Ought to Know About Race as a Variable.” *Health Services Research* 29(1):1–16.
- Lee, S.M. 2001. “Using the New Racial Categories in the 2000 Census.” KIDS COUNT/PRB Report. The Annie E. Casey Foundation and The Population Reference Bureau, Washington, DC.
- Liebler, C.A. 2001. “The Fringes of American Indian Identity.” Ph.D. dissertation. Department of Sociology, University of Wisconsin–Madison.
- . 2004. “Ties on the Fringes of Identity.” *Social Science Research* 33:702–23.
- Mays, V.M., N.A. Ponce, D.L. Washington, and S.D. Cochran. 2003. “Classification of Race and Ethnicity: Implications for Public Health.” *Annual Review of Public Health* 24:83–110.

- National Committee on Vital and Health Statistics. 2005. "Eliminating Health Disparities: Strengthening Data on Race, Ethnicity, and Primary Language in the United States." U.S. Department of Health and Human Services, Washington, DC. Available online at <http://www.cdc.gov/nchs/data/misc/EliHealthDisp.pdf>
- National Institutes of Health, Office of Extramural Research. 2001. "NIH Policy and Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research—Amended, October, 2001." Available online at [http://grants.nih.gov/grants/funding/women\\_min/guidelines\\_amended\\_10\\_2001.htm](http://grants.nih.gov/grants/funding/women_min/guidelines_amended_10_2001.htm)
- Nobles, M. 2000. *Shades of Citizenship: Race and Census in Modern Politics*. Stanford, CA: Stanford University Press.
- Office of Management and Budget (OMB). 1997. "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity." Federal Register 62FR58781-58790. Available online at <http://www.whitehouse.gov/omb/fedreg/1997standards.html>
- . 2000. "Provisional Guidance on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity." Available online at [http://www.whitehouse.gov/omb/inforeg/re\\_guidance2000update.pdf](http://www.whitehouse.gov/omb/inforeg/re_guidance2000update.pdf)
- Parker, J.D. and D.M. Makuc. 2002. "Methodologic Implications of Allocating Multiple-Race Data to Single-Race Categories." *Health Services Research* 37(1):201–13.
- Parker, J.D., N. Schenker, D.D. Ingram, J.A. Weed, K.E. Heck, and J.H. Madans. 2004. "Bridging Between Two Standards for Collecting Information on Race and Ethnicity: An Application to Census 2000 and Vital Rates." *Public Health Reports* 119:192–205.
- Renn, K.A. 2004. *Mixed Race Students in College: the Ecology of Race, Identity, and Community on Campus*. Albany, NY: SUNY Press.
- Renn, K.A. and C.J. Lunceford. 2004. "Because the Numbers Matter: Transforming Postsecondary Education Data on Student Race and Ethnicity to Meet the Challenges of a Changing Nation." *Educational Policy* 18:752–83.
- Rockquemore, K.A. 2002. "Negotiating the Color Line: The Gendered Process of Racial Identity Construction Among Black/White Biracial Women." *Gender and Society* 16:485–503.
- Rockquemore, K.A. and D.L. Brunsma. 2002. *Beyond Black: Biracial Identity in America*. Thousand Oaks, CA: Sage.
- Rodríguez, C.E. 1992. "Race, Culture, and Latino 'Otherness' in the 1980 Census." *Social Science Quarterly* 73:930–37.
- . 2000. *Changing Race: Latinos, the Census, and the History of Ethnicity in the United States*. New York: New York University Press.
- Root, M.P.P., ed. 1996. *The Multiracial Experience: Racial Borders as the New Frontier*. Newbury Park, CA: Sage.
- Schenker, N. and J.D. Parker. 2003. "From Single-Race Reporting to Multiple-Race Reporting: Using Imputation Methods to Bridge the Transition." *Statistics in Medicine* 22:1571–87.
- Snipp, C.M. 2003. "Racial Measurement in the American Census: Past Practices and Implications for the Future." *Annual Review of Sociology*. 29:563–88.
- Tucker, C.R., S. Miller, and J. Parker. 2002. "Comparing Census Race Data Under the Old and the New Standards." Pp. 365–90 in *The New Race Question: How the Census Counts Multiracial Individuals*, J. Perlmann and M.C. Waters, eds. New York: Russell Sage Foundation.
- Wallace, K.R. 2001. *Relative/Outsider: The Art and Politics of Identity Among Mixed Heritage Students*. Westport, CT: Ablex.
- Waters, M.C. 1999. *Black Identities: West Indian Immigrant Dreams and American Realities*. Cambridge, MA: Harvard University Press.
- . 2000. "Immigration, Intermarriage, and the Challenges of Measuring Racial/Ethnic Identities." *American Journal of Public Health* 90(11):1735–37.
- Xie, Y. and K. Goyette. 1998. "The Racial Identification of Biracial Children With One Asian Parent: Evidence From the 1990 Census." *Social Forces* 76:547–70.